

2590
1009

#2

OIPE

RAW SEQUENCE LISTING

DATE: 12/04/2001

PATENT APPLICATION: US/09/849,866

TIME: 12:10:23

Input Set : A:\GENSET.15CDV1.SEQ.txt

Output Set: N:\CRF3\11212001\I849866.raw

SEQUENCE LISTING

ENTERED

```

1 (1) GENERAL INFORMATION:
3   (i) APPLICANT: Ilya Chumakov
4             Hiroaki Tanaka
6   (ii) TITLE OF INVENTION: High Throughput DNA Sequencing Vector
8   (iii) NUMBER OF SEQUENCES: 22
C--> 10   (iv) CORRESPONDENCE ADDRESS:
11         (A) ADDRESSEE: Knobbe, Martens, Olson & Bear, LLP
12         (B) STREET: 550 West C Street, Suite 1200
13         (C) CITY: San Diego
C--> 14         (D) STATE: California
15         (E) COUNTRY: USA
16         (F) ZIP: 92101
18   (v) COMPUTER READABLE FORM:
19         (A) MEDIUM TYPE: Floppy Disk
20         (B) COMPUTER: IBM PC compatible
21         (C) OPERATING SYSTEM: Win95
22         (D) SOFTWARE: Word
C--> 0   (vi) CURRENT APPLICATION DATA:
C--> 0         (A) APPLICATION NUMBER: US/09/849,866
C--> 0         (B) FILING DATE: 04-May-2001
24   (viii) ATTORNEY/AGENT INFORMATION:
25         (A) NAME: Daniel Hart
26         (B) REGISTRATION NUMBER: 40,637
27         (C) REFERENCE/DOCKET NUMBER: GENSET.15CDV1
29   (ix) TELECOMMUNICATION INFORMATION:
30         (A) TELEPHONE: (619) 235-8550
31         (B) TELEFAX: (619) 235-0176
34 (2) INFORMATION FOR SEQ ID NO: 1:
36   (i) SEQUENCE CHARACTERISTICS:
37         (A) LENGTH: 10317 base pairs
38         (B) TYPE: NUCLEIC ACID
39         (C) STRANDEDNESS: DOUBLE
40         (D) TOPOLOGY: CIRCULAR
W--> 42   (ii) MOLECULE TYPE: synthetic DNA
44   (vi) ORIGINAL SOURCE:
45         (A) ORGANISM: Cloning vector pGenDEL
47   (ix) FEATURE:
48         (A) NAME/KEY: pGendel
49         (B) LOCATION: 1..10317
51   (ix) FEATURE:
52         (A) NAME/KEY: Homology with X06404 compl (411..1668)
53         (B) LOCATION: 9..1266
54         (C) IDENTIFICATION METHOD: blastn against X06404
56   (ix) FEATURE:
57         (A) NAME/KEY: Kanamycin resistance gene CDS
58         (B) LOCATION: 142..957

```

RAW SEQUENCE LISTING

DATE: 12/04/2001

PATENT APPLICATION: US/09/849,866

TIME: 12:10:23

Input Set : A:\GENSET.15CDV1.SEQ.txt

Output Set: N:\CRF3\11212001\I849866.raw

```

59         (C) IDENTIFICATION METHOD: By homology to X06404
60
61     (ix) FEATURE:
62         (A) NAME/KEY: Tn1000'sright end
63         (B) LOCATION: complement 1332..1371
64         (C) IDENTIFICATION METHOD: blastn against X60200)
65
66     (ix) FEATURE:
67         (A) NAME/KEY: Homology with U46017 (1-472)
68         (B) LOCATION: 1423..1894
69         (C) IDENTIFICATION METHOD: blastn against U46017
70
71     (ix) FEATURE:
72         (A) NAME/KEY: single stranded DNA replication origin
73         (B) LOCATION: 1423..1894
74         (C) IDENTIFICATION METHOD: By homology to U46017
75         (D) OTHER INFORMATION: mutation T -> C 1658
76
77     (ix) FEATURE:
78         (A) NAME/KEY: Homology with U51113 (2382..6997)
79         (B) LOCATION: 1896..6544
80         (C) IDENTIFICATION METHOD: blastn against U51113
81
82     (ix) FEATURE:
83         (A) NAME/KEY: OriS
84         (B) LOCATION: 1972..2188
85         (C) IDENTIFICATION METHOD: By homology to U51113
86
87     (ix) FEATURE:
88         (A) NAME/KEY: repELR
89         (B) LOCATION: 2897..2918
90         (D) OTHER INFORMATION: Described in seqID 16
91
92     (ix) FEATURE:
93         (A) NAME/KEY: RepE
94         (B) LOCATION: 2903..3034
95         (C) IDENTIFICATION METHOD: By homology to U51113
96
97     (ix) FEATURE:
98         (A) NAME/KEY: T3
99         (B) LOCATION: 3043..3059
100        (D) OTHER INFORMATION: Described in seqID 17
101
102     (ix) FEATURE:
103         (A) NAME/KEY: LRT3RA
104         (B) LOCATION: complement 3045..3069
105         (D) OTHER INFORMATION: Described in seqID 15
106
107     (ix) FEATURE:
108         (A) NAME/KEY: IncC
109         (B) LOCATION: 3070..3320
110         (C) IDENTIFICATION METHOD: By homology to U51113
111         (D) OTHER INFORMATION: insertion 33 bases 3038..3071
112
113     (ix) FEATURE:
114         (A) NAME/KEY: ParA
115         (B) LOCATION: 3655..4821
116         (C) IDENTIFICATION METHOD: By homology to U51113
117         (D) OTHER INFORMATION: mutation G -> A 3878
118
119     (ix) FEATURE:

```

RAW SEQUENCE LISTING

DATE: 12/04/2001

PATENT APPLICATION: US/09/849,866

TIME: 12:10:23

Input Set : A:\GENSET.15CDV1.SEQ.txt

Output Set: N:\CRF3\11212001\I849866.raw

```

120      (A) NAME/KEY: ParB
121      (B) LOCATION: 4821..5792
122      (C) IDENTIFICATION METHOD: By homology to U51113
124      (ix) FEATURE:
125          (A) NAME/KEY: ParC
126          (B) LOCATION: 5865..6382
127          (C) IDENTIFICATION METHOD: By homology to U51113
129      (ix) FEATURE:
130          (A) NAME/KEY: Homology with J01688 (complement 175..819)
131          (B) LOCATION: 6574..7218
132          (C) IDENTIFICATION METHOD: blastn against J01688
133          (D) OTHER INFORMATION: mutation A -> G 7096
135      (ix) FEATURE:
136          (A) NAME/KEY: CDS streptomycin sensitivity gene
137          (B) LOCATION: complement 6716..7090
138          (C) IDENTIFICATION METHOD: By homology to J01688
139          (D) OTHER INFORMATION: mutation A -> G 6728
140      mutation G -> C 6821
141      mutation C -> T 6866
142      mutation T -> C 7013
143      mutation T -> A 7058
145      (ix) FEATURE:
146          (A) NAME/KEY: rpsLR
147          (B) LOCATION: 7155..7174
148          (D) OTHER INFORMATION: Described in seqID 12
150      (ix) FEATURE:
151          (A) NAME/KEY: SP6
152          (B) LOCATION: 7230..7248
153          (D) OTHER INFORMATION: Described in seqID 13
155      (ix) FEATURE:
156          (A) NAME/KEY: Tn1000's left end
157          (B) LOCATION: 7252..7291
158          (C) IDENTIFICATION METHOD: blast (X60200)
160      (ix) FEATURE:
161          (A) NAME/KEY: Homology with X02730 (complement 37..1959)
162          (B) LOCATION: 7305..9227
163          (C) IDENTIFICATION METHOD: blastn against X02730
165      (ix) FEATURE:
166          (A) NAME/KEY: CDS levansucrase gene
167          (B) LOCATION: complement 7379..8800
168          (C) IDENTIFICATION METHOD: By homology to X02730
169          (D) OTHER INFORMATION: mutation T -> C 7466
170      mutation A -> G 7739
171      mutation T -> C (Asn -> Asp) 8347
172      mutation T -> C 8600
173      mutation G -> A (Ala -> Val) 8772
177      (ix) FEATURE:
178          (A) NAME/KEY: SLR3
179          (B) LOCATION: 8711..8731

```

RAW SEQUENCE LISTING

DATE: 12/04/2001

PATENT APPLICATION: US/09/849,866

TIME: 12:10:23

Input Set : A:\GENSET.15CDV1.SEQ.txt

Output Set: N:\CRF3\11212001\I849866.raw

```

180      (D) OTHER INFORMATION: Described in seqID 14
182  (ix) FEATURE:
183      (A) NAME/KEY: Homology with J01636 (complement 1158..1465)
184      (B) LOCATION: 9298..9623
185      (C) IDENTIFICATION METHOD: blastn against J01636
187  (ix) FEATURE:
188      (A) NAME/KEY: CDS alpha peptide beta-galactosidase
189      (B) LOCATION: complement 9276..9497
190      (C) IDENTIFICATION METHOD: By homology to J01636
192  (ix) FEATURE:
193      (A) NAME/KEY: primer HE1
194      (B) LOCATION: complement 9465..9479
196  (ix) FEATURE:
197      (A) NAME/KEY: primer HE2
198      (B) LOCATION: 9461..9475
200  (ix) FEATURE:
201      (A) NAME/KEY: primer LacLRS2Avr
202      (B) LOCATION: complement 9603..9630
204  (ix) FEATURE:
205      (A) NAME/KEY: primer LacE2Mlu
206      (B) LOCATION: 9289..9314
208  (ix) FEATURE:
209      (A) NAME/KEY: Homology with M77789 (1889..2576)
210      (B) LOCATION: 9629..10315
211      (C) IDENTIFICATION METHOD: blastn against M77789
213  (ix) FEATURE:
214      (A) NAME/KEY: high copy-number double-stranded DNA replication origin
215      (B) LOCATION: complement 9629..10315
216      (C) IDENTIFICATION METHOD: By homology to M77789
217      (D) OTHER INFORMATION: mutation C -> T 9803
218 site ScaI 10029 - 10034
219 site PmlI 10038 - 10043
220 CLONING SITES 10031 - 10041
223  (ix) FEATURE:
224      (A) NAME/KEY: oriLRd
225      (B) LOCATION: 9856..9881
226      (D) OTHER INFORMATION: Described in seqID 8
228  (ix) FEATURE:
229      (A) NAME/KEY: OS1
230      (B) LOCATION: 10009..10026
231      (D) OTHER INFORMATION: Described in seqID 10
233  (ix) FEATURE:
234      (A) NAME/KEY: OR1
235      (B) LOCATION: complement 10046..10062
236      (D) OTHER INFORMATION: Described in seqID 11
238  (ix) FEATURE:
239      (A) NAME/KEY: oriLRr
240      (B) LOCATION: complement 10182..10202
241      (D) OTHER INFORMATION: Described in seqID 9

```

RAW SEQUENCE LISTING

DATE: 12/04/2001

PATENT APPLICATION: US/09/849,866

TIME: 12:10:24

Input Set : A:\GENSET.15CDV1.SEQ.txt

Output Set: N:\CRF3\11212001\I849866.raw

```

243      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
246 GACCGTTTGT CGACCTGCAG GGGGGGGGGG GAAAGCCACG TTGTGTCTCA AAATCTCTGA      60
248 TGTTACATTG CACAAGATAA AAATATATCA TCATGAACAA TAAAACTGTC TGCTTACATA      120
250 AACAGTAATA CAAGGGGTGT TATGAGCCAT ATTCAACGGG AAACGTCTTG CTCGAGGCCG      180
252 CGATTAAATT CCAACATGGA TGCTGATTTA TATGGGTATA AATGGGCTCG CGATAATGTC      240
254 GGGCAATCAG GTGCGACAAT CTATCGATTG TATGGGAAGC CCGATGCGCC AGAGTTGTTT      300
256 CTGAAACATG GCAAAGGTAG CGTTGCCAAT GATGTTACAG ATGAGATGGT CAGACTAAAC      360
258 TGGCTGACGG AATTTATGCC TCTTCCGACC ATCAAGCATT TTATCCGTAC TCCTGATGAT      420
260 GCATGGTTAC TCACCACTGC GATCCCCGGG AAAACAGCAT TCCAGGTATT AGAAGAATAT      480
262 CCTGATTGAG GTGAAAATAT TGTTGATGCG CTGGCAGTGT TCCTGCGCCG GTTGCATTCT      540
264 ATTCCTGTTT GTAATTGTCC TTTTAACAGC GATCGCGTAT TTCGTCTCGC TCAGGCGCAA      600
266 TCACGAATGA ATAACGGTTT GGTGATGCG AGTGATTTTG ATGACGAGCG TAATGGCTGG      660
268 CCTGTTGAAC AAGTCTGGAA AGAAATGCAT AAGCTTTTGC CATTCTCACC GGATTCAGTC      720
270 GTCACATCAT GTGATTTCTC ACTTGATAAC CTTATTTTGT ACGAGGGGAA ATTAATAGGT      780
272 TGTATTGATG TTGGACGAGT CGGAATCGCA GACCGATACC AGGATCTTGC CATCCTATGG      840
274 AACTGCCTCG GTGAGTTTTT TCCTTCATTA CAGAAACGGC TTTTTCAAAA ATATGGTATT      900
276 GATAATCCTG ATATGAATAA ATTGCAGTTT CATTGATGTC TCGATGAGTT TTTCTAATCA      960
278 GAATTGGTTA ATTGGTTGTA ACACTGGCAG AGCATTACGC TGACTTGACG GGACGGCGGC      1020
280 TTTGTTGAAT AAATCGAACT TTTGCTGAGT TGAAGGATCA GATCACGCAT CTTCCCGACA      1080
282 ACGCAGACCG TTCCGTGGCA AAGCAAAAGT TCAAAATCAC CAACTGGTCC ACCTACAACA      1140
284 AAGCTCTCAT CAACCGTGGC TCCCTCACTT TCTGGCTGGA TGATGGGGCG ATTCAGGCCT      1200
286 GGTATGAGTC AGCAACACCT TCTTCACGAG GCAGACCTCA GCGCCCCCCC CCCCCTGCAG      1260
288 GTCGACTATA CAACGATCCC GCCATACCA GGCCATCTGG CTGGGGTGCT TAACCGTAAG      1320
290 TCTGACGAAT TGGGGTTTGA GGGCCAATGG AACGAAAACG TACGTAAAGG ATCAGTTCCC      1380
292 TATAGTGAGT CGTATTAGCG GCCAGATCGA TCTAAGTGCC ACCTAAATTG TAAGCGTTAA      1440
294 TATTTTGTGA AAATTCGCGT TAAATTTTTG TTAAATCAGC TCATTTTTTA ACCAATAGGC      1500
296 CGAAATCGGC AAAATCCCTT ATAAATCAAA AGAATAGACC GAGATAGGGT TGAGTGTTGT      1560
298 TCCAGTTTGG AACAAGAGTC CACTATTAAA GAACGTGGAC TCCAACGTCA AAGGGCGAAA      1620
300 AACCGTCTAT CAGGGCGATG GCCCCTACG TGAACCAACA CCCTAATCAA GTTTTTTGGG      1680
302 GTCGAGGTGC CGTAAAGCAC TAAATCGGAA CCCTAAAGGG AGCCCCCGAT TTAGAGCTTG      1740
304 ACGGGGAAAG CCGGCGAACG TGGCGAGAAA GGAAGGGAAG AAAGCGAAAG GAGCGGGCGC      1800
306 TAGGGCGCTG GCAAGTGTAG CGGTCACGCT GCGCGTAACC ACCACACCCG CCGCGCTTAA      1860
308 TGCGCCGCTA CAGGGCGCGT CCCATTCGCC ATTCGTCGAG TGAGCGAGGA AGCACCAGGG      1920
310 AACAGCACTT ATATATTCTG CTTACACACG ATGCCTGAAA AAACCTCCCT TGGGGTTATC      1980
312 CACTTATCCA CGGGGATATT TTTATAATTA TTTTTTTTAT AGTTTTTATA TCTTCTTTTT      2040
314 TAGAGCGCCT TGTAGGCCTT TATCCATGCT GGTCTAGAG AAGGTGTTGT GACAAATTGC      2100
316 CCTTTCAGTG TGACAAATCA CCCTCAAATG ACAGTCCTGT CTGTGACAAA TTGCCCTTAA      2160
318 CCCTGTGACA AATTGCCCTC AGAAGAAGCT GTTTTTTTCAC AAAGTTATCC CTGCTTATTG      2220
320 ACTCTTTTTT ATTTAGTGTG ACAATCTAAA AACTTGTCAC ACTTCACATG GATCTGTCAT      2280
322 GCGGAAACA GCGGTTATCA ATCACAAGAA ACGTAAAAAT AGCCCGCGAA TCGTCCAGTC      2340
324 AAACGACCTC ACTGAGGCGG CATATAGTCT CTCCCGGGAT CAAAAACGTA TGCTGTATCT      2400
326 GTTCGTTGAC CAGATCAGAA AATCTGATGG CACCCTACAG GAACATGACG GTATCTGCGA      2460
328 GATCCATGTT GCTAAATATG CTGAAATATT CGGATTGACC TCTGCGGAAG CCAGTAAGGA      2520
330 TATACGGCAG GCATTGAAGA GTTTCGCGGG GAAGGAAGTG GTTTTTTATC GCCCTGAAGA      2580
332 GGATGCCGGC GATGAAAAAG GCTATGAATC TTTTCCTTGG TTTATCAAAC GTGCGCACAG      2640
334 TCCATCCAGA GGGCTTTACA GTGTACATAT CAACCCATAT CTCATTCCCT TCTTTATCGG      2700
336 GTTACAGAAC CGGTTTACGC AGTTTCGGCT TAGTGAAACA AAAGAAATCA CCAATCCGTA      2760
338 TGCCATGCGT TTATACGAAT CCCTGTGTCA GTATCGTAAG CCGGATGGCT CAGGCATCGT      2820
340 CTCTCTGAAA ATCGACTGGA TCATAGAGCG TTACCAGCTG CCTCAAAGTT ACCAGCGTAT      2880

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/849,866

DATE: 12/04/2001

TIME: 12:10:25

Input Set : A:\GENSET.15CDV1.SEQ.txt

Output Set: N:\CRF3\11212001\I849866.raw

L:10 M:220 C: Keyword misspelled or invalid format, [(iv) CORRESPONDENCE ADDRESS:]
L:14 M:220 C: Keyword misspelled or invalid format, [(D) STATE:]
L:0 M:249 C: Inserted Mandatory Field, [(vi) CURRENT APPLICATION DATA:]
L:0 M:249 C: Inserted Mandatory Field, [(A) APPLICATION NUMBER:]
L:0 M:249 C: Inserted Mandatory Field, [(B) FILING DATE:]
L:42 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=1
L:600 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=2
L:622 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=3
L:644 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=4
L:666 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=5
L:688 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=6
L:710 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=7
L:754 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=8
L:781 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:776 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=9
L:798 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=10
L:820 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=11
L:842 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=12
L:864 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=13
L:886 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=14
L:908 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=15
L:930 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=16
L:952 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=17
L:974 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=18
L:1021 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=19
L:1068 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=20
L:1090 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=21
L:1111 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=22